Relation entre transitivité et structure de communauté dans les réseaux complexes

Keziban Gunce Orman[*,**] Vincent Labatut[*], Hocine Cherifi[**]

[*]Galatasaray University, Computer Science Department, Istanbul, Turkey
[**]University of Burgundy, LE2I UMR CNRS 5158, Dijon, France
korman@gsu.edu.tr

Résumé. Dans les réseaux complexes, les relations entre la structure de communauté et d'autres propriétés topologiques n'a pas encore été analysées clairement. Dans cette étude préliminaire, nous explorons expérimentalement le lien potentiel entre transitivité et structure de communauté. Sur des réseaux générés artificiellement, nous étudions l'évolution de la transitivité en fonction du niveau de definition de la structure communautaire. Puis, nous observons l'évolution de la structure de communauté sur des réseaux dont nous contrôlons la transitivité. Les résultats obtenus suggèrent qu'il existe une relation entre transitivité élevée et structure de communauté. Un travail analytique complémentaire est nécessaire pour identifier la nature exacte de cette relation.

1 Introduction

Dans un réseau complexe, une communauté est un sous-ensemble de nœuds plus densément interconnectés,relativement au reste du réseau. Différentes définitions formelles de ce concept existent, généralement basées sur la comparaison des nombres de liens inter et intra-communautaires. Celle implémentée par la mesure de modularité de Newman est certainement la plus connue et la plus utilisée. On dit qu'un réseau possède une structure de communauté quand il prend la forme d'un ensemble de communautés interconnectées. De nombreux systèmes du monde réel sont modélisables sous forme de réseaux complexes, et ceux-ci possèdent communément une structure de communauté. On peut supposer que la présence d'une telle structure est liée à d'autres propriétés topologiques mesurables.Pourtant, ce type de relation n'a pas été clairement examiné dans les travaux existants. Dans cette étude, nous nous concentrons sur la relation potentielle entre transitivité et structure de communauté. La transitivité d'un réseau (aussi appelée coefficient de clustering) correspond à la proportion de triangles (groupes de trois nœuds connectés) parmi toutes les triades connectées (groupes de trois nœuds dont au moins deux connectés). La relation peut paraître triviale au premier abord, mais il est cependant aisé d'y trouver des contre-exemples : un réseau peut être fortement transitif sans avoir pour autant une structure de communauté (ex. : réseau complètement connecté), et il peut exhiber une structure de communauté sans être transitif (ex. : communautés multiparties). Ici, nous choisissons de faire abstraction de ce type de construction artificielle et de nous concentrer sur des réseaux aux propriétés plus réalistes, i.e. proches de celles observés dans des modèles de systèmes réels. Pour ce faire, nous réalisons deux phases de tests. Lors de la première, nous utilisons un modèle permettant de générer artificiellement des réseaux dont le niveau de séparation des communautés est contrôlé, afin d'analyser comment un tel changement affecte la transitivité. Lors de la seconde phase, nous étudions comment l'évolution de la transitivité affecte la structure de communauté. Nous utilisons deux modèles, capables de créer des réseaux avec une transitivité élevée. Afin d'identifier la structure des réseaux, nous utilisons les algorithmes de détection de communautés Louvain (Blondel et al., 2008) et Infomap (Rosvall et Bergstrom, 2008). Le niveau de séparation des structures de communauté obtenues est quantifié au moyen de la modularité. Les sections 2 et 3 présentent nos résultats pour ces deux phases expérimentales. Nous concluons ensuite avec les limitations de notre travail et les points que nous comptons aborder par la suite.

2 Modèle à structure de communauté

Afin d'observer la relation entre structure de communauté et transitivité, nous avons utilisé le modèle LFR de Lancichinetti *et al.* (Lancichinetti et al., 2008). Il est capable de générer des réseaux dont le degré des nœuds et la taille des communautés suivent tous les deux une loi de puissance, ce qui est considéré comme un comportement réaliste. La structure de communauté est contrôlée grâce à un paramètre appelé *coefficient de mélange* et noté μ. Celui-ci indique la proportion moyenne de liens externes possédés par un nœud, i.e. le connectant à des nœuds extérieurs à sa propre communauté. La générat on se déroule en trois étapes. Premièrement, le modèle de configuration (CM) (Bender et Canfield, 1978) est utilisé pour générer un réseau initial respectant la distribution de degré voulue. Deuxièmement, des communautés virtuelles sont définies de manière à suivre la distribution de tailles prédéfinie. Troisièmement, un processus itératif permet de reconnecter certains liens, de manière à approcher μ sans pour autant modifier la distribution du degré. Afin de nous affranchir d'un éventuel effet du modèle initial, nous avons modifié le modèle LFR original, en substituant différents modèles lors de la première étape. Nous avons utilisé non seulement CM, mais aussi le modèle Barabási–Albert (BA) (Barabasi et Albert, 1999) et un modèle évolutionnaire (EV) (Poncela et al., 2008). En utilisant ces trois variantes, nous avons généré puis analysé trois collections différentes de réseaux de taille 5000, correspondant aux trois modèles initiaux testés : LFR- CM, LFR-BA et LFR-EV. Au vue de nos résultats, la première observation que nous pouvons faire est que l'introduction d'une structure de communauté dans un réseau (grâce à l'étape de reconnexion de LFR) fait augmenter sa transitivité de manière extrêmement importante. On observe des différences entre les collections vis-à-vis de ce phénomène : CM atteint les valeurs les plus élevées, suivi par EV puis BA. Cependant, la différence par rapport à la transitivité mesurée sur les réseaux initiaux est non-négligeable pour les trois collections. La deuxième observation est que cette augmentation est directement liée au coefficient de mélange μ , comme l'illustre la figure 1 (*gauche*). La transitivité maximale est atteinte pour des valeurs proches de $\mu = 0$ (communautés très clairement séparées) et diminue lorsqu'on considère un μ proche de 1 (communautés difficiles à discerner). Là encore, ceci est valable pour les trois collections. Ces deux observations semblent indiquer que la présence d'une structure de communauté implique un certain niveau de transitivité.

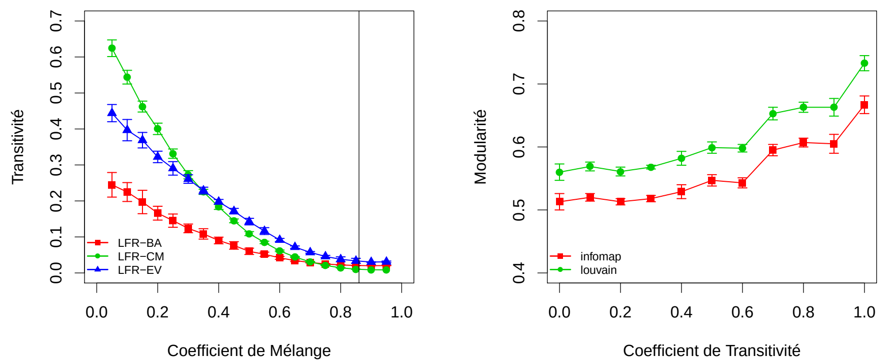

FIG. 1 – *Influence de μ sur la transitivité (gauche) : pour chaque valeur de μ, 25 réseaux ont été générés avec les paramètres $n = 5000, \langle k \rangle = 30, \gamma = 3, \beta = 2$ en appliquant le modèle LFR. Influence de la transitivité sur la modularité (droit) : pour chaque valeur du coefficient de transitivité, 6 réseaux de taille $n = 1000$ ont été générés avec le modèle de Newman.*

3 Modèles transitifs

Pour étudier l'effet de la transitivité sur la structure de communauté, nous avons utilisé deux modèles différents permettant de générer des réseaux avec une transitivité élevée et/ou contrôlée. Le premier, dû à Newman (Newman, 2009), est une modification du CM permettant de contrôler la transitivité des réseaux. Dans ce modèle, le degré k_i d'un nœud i est $k_i = s_i + 2t_i$ tel que t_i est le nombre de triangles auquel il appartient et s_i est les liens simples (n'appartenant pas à des triangles). Dans notre implémentation, la transitivité est contrôlée au moyen d'un paramètre appelé *coefficient de transitivité*. L'une des limitations de ce modèle est que la structure transitive qu'il génère n'est pas forcément réaliste. Par exemple, les triangles générés n'auront quasiment jamais plus d'un seul nœud commun, et ne partageront donc pas de côté. Nous avons mis au point un modèle qui permet de pallier cet inconvénient. Celui-ci prend en entrée la séquence de degrés voulue. Dans un premier temps, un réseau en anneau est généré pour puis, les liens manquants sont rajoutés un par un, de façon aléatoire mais contrainte. D'abord, la séquence de degrés voulue doit être respectée. De plus, nous favorisons la création de liens entre des nœuds voisins afin d'accroitre la transitivité. Après avoir généré des réseaux en utilisant chacun de ces deux modèles, nous avons appliqué deux algorithmes de détection de communautés récents : Louvain et Infomap. Nous avons évalué l'existence d'une structure de communauté en considérant le niveau de modularité atteint par ces algorithmes. Louvain optimise directement cette mesure, alors qu'Infomap utilise une approche complètement différente basée sur la théorie de l'information. Ceci rend ces deux algorithmes complémentaires. Nous avons d'abord généré 25 réseaux de taille $n = 5000$ et de degré moyen $k_i = 30$ avec le modèle que nous avons proposé. Louvain et Infomap atteignent tous les deux une modularité de 0.8 pour une transitivité de 0.3 en moyenne. Le modèle de Newman permet de contrôler la transitivité de façon relativement fine. Nous avons donc généré des échantillons de 6 réseaux avec les paramètres $n = 1000$ et $k_i = 5$, et en utilisant des valeurs croissantes du coefficient de transitivité. Comme illustré dans la figure 1 (*droit*), la modularité croit avec le coefficient de transitivité, notamment pour les valeurs supérieures à 0.4.

4 Conclusion

Dans ce travail, nous nous sommes intéressés à la relation entre la structure de communauté et la transitivité des réseaux complexes. Nous avons tout d'abord utilisé le modèle LFR pour générer des réseaux artificiels réalistes possédant une structure de communauté plus ou moins forte. Nous avons observé qu'une modification de la structure de communauté est toujours associée à un changement de transitivité : plus les communautés sont séparées et plus la transitivité est élevée (tous les autres paramètres étant fixés). À l'aide d'un modèle défini par Newman et d'un autre développé par nos soins, nous avons ensuite généré des réseaux à transitivité élevée et/ou contrôlée. Dans les deux cas, nous avons observé que la modularité, mesurée avec deux algorithmes de détection de communauté différents (Louvain et Infomap) augmente avec leur transitivité. Ces résultats expérimentaux indiquent qu'il existe une relation entre la transitivité et la structure de la communauté dans ce type de réseaux. Cependant, les résultats préliminaires de cette étude sont limités et demandent à être confirmés et généralisés par des résultats analytiques.

Références

Barabasi, A.-L. et R. Albert (1999). Emergence of scaling in random networks. *Science 286*, 509–512.

Bender, E. et E. R. Canfield (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory - Series A 24*, 296–307.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, et E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment 2008*.

Lancichinetti, A., S. Fortunato, et F. Radicchi (2008). Benchmark graphs for testing community detection algorithms. *Phys. Rev. E 78*, 046110.

Newman, M. (2009). Random graphs with clustering. *Physical Review Letters 103*.

Poncela, J., J. Gomez-Gardenes, L. M. Floria, A. Sanchez, et Y. Moreno (2008). Complex cooperative networks from evolutionary preferential attachment. *PLoS ONE 3*, e2449.

Rosvall, M. et C. Bergstrom (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences 105*.

Summary

The relation between community structure and other topological properties in complex networks, has not been analyzed clearly. In this preliminary study, we explore experimentally the potential link between transitivity and community structure. We study the evolution of transitivity based on the quality of the community structure on artificially generated networks. Then we observe the evolution of community structure according to transitivity value. The results show a relationship between high transitivity and community structure. Additional analytical work is needed to identify the exact nature of this relationship.